

IN-82
053-76

Representation and Integration of Scientific Information

NASA Interchange for Joint Research

Award NCC 2-~~5347~~
5247

FINAL REPORT

The objective of this Joint Research Interchange with NASA-Ames was to investigate how the Tsimmis technology could be used to represent and integrate scientific information. The funding allowed us to work with researchers within NAS at the NASA Ames Research Center, to understand their information needs, and to work with them on integration strategies.

Most organizations have a need to access and integrate information from multiple, disparate information sources that may include both structured as well as semi-structured information. At Stanford we have been working on an information integration project called Tsimmis, supported by DARPA. The main goal of the Tsimmis project is to allow a decision maker to find information of interest from such sources, fuse it, and process it (e.g., summarize it, visualize it, discover trends). Another important goal is the easy incorporation of new sources, as well the ability of deal with sources whose structure or services evolve.

During the Interchange we had research meetings approximately every month or two. The participants from NASA included Michael Cox and Peter Vanderbilt. The Stanford PI, and various students and Stanford staff members also participated. NASA researchers also participated in some of our regular Tsimmis meetings. As planned, our meetings discussed problems and solutions to various information integration problems.

The funds provided by NASA supported work that lead to the following two papers:

- (1) R. Yerneni, Y. Papakonstantinou, S. Abiteboul, H. Garcia-Molina
"Fusion Queries over Internet Databases,"
Proceedings of the 6th International Conference on
Extending Database Technology,
(Springer Verlag LNCS 1377), Valencia, Spain, March 1998, pp. 57-71.
Available at: <http://www-db.stanford.edu/pub/papers/fusion-edbt.ps>

This paper studies alternatives for efficiently executing fusion queries. Information fusion was identified in our original proposal as one of the problems of common interest. This paper considers queries that fuse information from several sources, and presents various schemes for optimizing them.

- (2) Orkut Buyukkokten, Arturo Crespo, H. Garcia-Molina,
"Efficient Query Subscription Processing in a Multicast Environment,"
submitted for publication, 1999.
Available at: <http://www-db.stanford.edu/pub/papers/badd.ps>

rec'd.
MAY 04 1999

CC: 202A-3

CASI

In a shared scientific information environment, dissemination of new information is important. This paper studies strategies for automatically multicasting new information to a set of subscribers. Algorithms are presented for merging the subscriptions of users with common needs, to improve efficiency. the algorithms are compared analytically and via detailed simulations.

No inventions resulted from this research.